Single-lab Pilot Study Design Progress Report

Background

A rough idea of an initial evaluation using existing data followed by the pilot study was developed. However, before a pilot study plan can be developed in any detail, guidance is needed on a key question:

a) Are the procedures intended to identify the absolute detection and quantitation limits achievable by the laboratory, or
b) Are they intended to provide a demonstration that the detection and quantitation limits claimed by the laboratory are as high as or higher than the absolute detection limits?

The answer may affect the design of the Pilot study.

The amount of work required on the part of the laboratory for (a) is much greater that that required for (b). Given that the absolute detection limit on a given instrument can vary quite a bit from day to day (or even sample to sample, especially for samples in a dirty matrix) the attempt to demonstrate an absolute detection limit may be close to impossible.

The detection and quantitation limit procedures to be evaluated vary in their approach to this question. The single lab part of the LCMRL and Ken Osborn's procedure lean towards a demonstration that the claimed laboratory limits are justifiable. The ACIL and Consensus group procedures lean towards determination of the absolute limits. Depending on the answer to the question, some of the procedures may need to be modified. If the FAC decides that both objectives may be important than all the procedures may need some modification.

Evaluation of Existing Data

There is a considerable amount of existing data that can be used to evaluate the effectiveness of some of the produced detection / quantitation procedures - this needs to be evaluated before embarking on a large scale pilot study in order to reduce costs. All existing data will need to go through a pre-qualification process determined by the Technical Work Group. Existing data may also reduce the range of tests that the pilot study needs to evaluate. Some data sources identified include:

a. Canadian lab data – John Phillips
b. Judy Morgan data – John Phillips
c. CLP data – Richard Burrows
d. Ken Osborn, Andy Eaton and Dave Kimborough data – Ken Osborn
e. USGS data – Richard Burrows
f. WRC data – Cliff Kirchmer
g. Tim Fitzpatrick's data

EPA has indicated that contactor assistance will be available to assist with this evaluation. Initially the data should be put into a standard format, this can be quite simple:

Expected concentration (Spike value) (Zero if sample is a method blank)	Result	Units	Date

Once in standard format, detection limits can be calculated using the first part of each data set, then Type 1 and Type II error rates assessed using the rest of the data set. The details of this assessment need to be developed by the TWG and approved by the FAC.

In general, much more data is available to evaluate uncensored Lc (method blanks) than censored Lc or Ld (low level spikes).

Objectives and Study Considerations

To determine if the proposed detection and quantitation limit procedures are workable and generate detection and quantitation limits that meet the needs of the environmental testing community in terms of the type I and type II error rates obtained (Lc and Ld) or precision and bias (Lq).

Study Considerations

1. Method types to be evaluated
   The detection / quantitation procedures need to be amenable to a number of different types of data. These include spectroscopic, GC/selective detector, GC/MS, SIM, IC, multi component analyte, gravimetric and others. In general it would be a good idea for the pilot to utilize methods that are approved at Part 136, but this is not so important for evaluation of existing data. For evaluation of procedures using existing data prior to the pilot, we propose using as many methods as we can find good data sets for. For the pilot, in order to reduce costs, only one method from each group a-f may be needed.

   a. Multi analyte metals
      i. ICP-OES (200.7 or 6010 or SM3120)
      ii. ICP/MS (200.8 or 6020 or SM3125)
   b. Single analyte metals
      i. 1631 mercury, any GFAA method
   c. Multi analyte chromatographic
      i. GC/MS volatiles (524.2, 624 or 8260 or SM 6200)
      ii. GC/MS semivolatiles (525, 625 or 8270 or SM6410)
      iii. GC/MS SIM PAH (modified 625 or 8270)
      iv. Organochlorine pesticides (508, 608 or 8081 or SM6630)
      v. Ion chromatography (300.0)

vi. PID (502, 602 or 8021)

d. Multi component analytes (Aroclors, Toxaphene)

i. 508, 608 or 8082 or SM6431

e. Gravimetric methods

i. HEM and SGT-HEM method 1664

f. General chemistry spectrophotometric methods

i. Cyanide (335.X)

ii. Phosphorous (365.X)

g. Others?

2. Is the amount of work required on the part of the laboratory reasonable? Note that moderate to large laboratories will need to perform several hundred detection limit studies (counting multi analyte methods as one study).

    a. This question was not discussed in any significant detail, but we did decide that either the initial spike concentrations used for calculation of the detection limits must be selected by the individual labs following instructions in the procedures, and cannot be a standard concentration used for all labs, or a wide range of spike concentrations must be used.

3. Can a detection limit be identified in a reasonable amount of time when analysis for a new method or analyte is required to be implemented at the laboratory?

    a. Some detection limit procedures have allowance for a short term startup procedure, others may need to have this added

4. Does the procedure identify Lc, Ld and/or Lq with a reasonable degree of accuracy for the method types listed above? In other words, are the type 1 and type 2 errors for work performed subsequent to the determination of the detection limit close to the targeted levels?

    a. Ideally at least 3 labs would test each procedure. We discussed the possibility of having one set of instructions that would develop data suitable for evaluation of the ACIL, consensus, Osborn MDL and LT-MDL (also possibly the ESI {Environmental Science Corporation, Judith Morgan} and at least part of the WRC) procedures. This limits some of our ability to the evaluate of the difficulty of each procedure but would be considerably less expensive and might set the stage for a final procedure that combined the best points from each. We will add the other procedures (MDL, LT-MDL, EBMUD) to the document that compares the salient features of the ACIL, Consensus and Judy Morgan (ESI) procedures so that the similarities and differences can be more easily evaluated.

    b. We hope that the single lab Hubbaux Vos procedure might be able to be evaluated as a subset of the multi lab evaluation.

    c. We thought that this determination of the accuracy of Lc for uncensored methods is fairly straightforward. Blanks subsequent to the critical value determination would be monitored and the number above LC evaluated. Determination of the accuracy of Lc for censored methods is difficult and the group considered that Curries

recommendation that Lc for these methods = zero was worth considering. Other options are included in the ACIL and Consensus group procedures. Determination of the accuracy of the Ld determination is fairly difficult for both types. The group thought that the best approach for evaluating the performance of the procedures would be to use blind spikes at a variety of concentrations encompassing the range of detection limits determined by the laboratories testing the procedure. If the range of detection limits was 4-10 then spikes at 2, 4, 6, 8, 10, 12, 16, 20, 24, and 30 might be used for determination of Ld and Lq.